

# Model-Based Co-clustering for Ordinal Data

Julien JACQUES<sup>1</sup> & Christophe BIERNACKI<sup>2</sup>

<sup>1</sup> Université de Lyon, Lumière Lyon 2, ERIC

<sup>2</sup> Université Lille 1 et Inria

# Ordinal data ?

---

## Definition

An **ordinal** variable  $\mu$  takes values among  $m$  **full ordered** modalities

$$\mu \in \{1, \dots, m\} \text{ with } 1 < \dots < m$$



## Widespread data

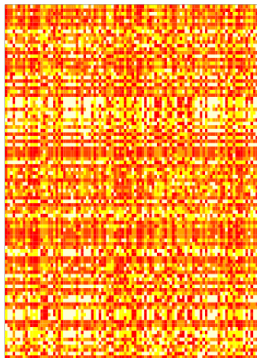
- Marketing: customer satisfaction surveys
- Sociology: education levels
- Medicine: pain evaluation
- **Many nominal data are . . . ordinal!**

# Co-clustering ?

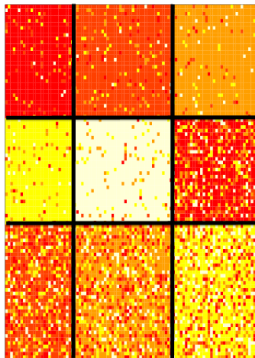
---

Simultaneous clustering of rows (individuals) and column (features)

original data



codustering result



# Overview

---

The BOS model for ordinal data

The Latent Block Model

Model inference

Numerical experiments

Model parameter estimation accuracy

Ability of ICL-BIC to select the number of co-clusters

Comparison with competitors

# Plan

---

The BOS model for ordinal data

The Latent Block Model

Model inference

Numerical experiments

Model parameter estimation accuracy

Ability of ICL-BIC to select the number of co-clusters

Comparison with competitors

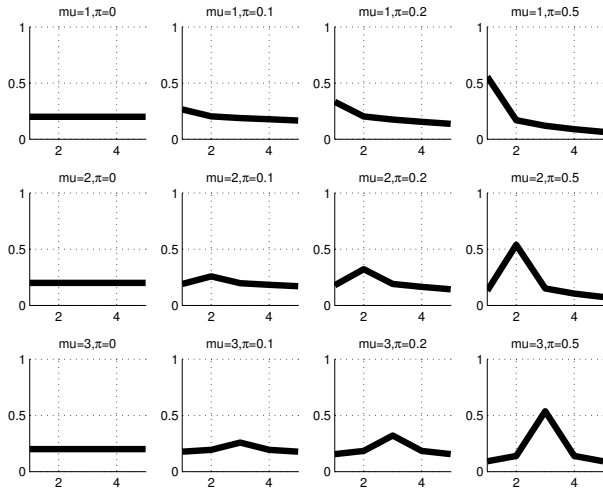
## BOS( $\mu, \pi$ ) model: parameters and properties [1]

---

- $\mu$ : **position** parameter (unique mode if  $\pi > 0$ )
- monotonic decrease around  $\mu$
- $\pi$ : **precision** parameter:
  - $p(\mu; \mu, \pi)$  increases with  $\pi$
  - $p(\mu; \mu, \pi) - p(x; \mu, \pi)$  increases with  $\pi$  ( $x \neq \mu$ )
  - uniform distribution if  $\pi = 0$
  - Dirac in  $\mu$  if  $\pi = 1$
- identifiability (if  $\pi = 0$ )

[1] *Biernacki & Jacques (2015), Model-based clustering of multivariate ordinal data relying on a stochastic binary search algorithm, to appear in Statistics and Computing.*

# BOS( $\mu, \pi$ ) model: illustration



## BOS( $\mu, \pi$ ) model: inference

---

Due to the nature of the BOS( $\mu, \pi$ ) model, maximum likelihood should be estimated thanks to an [EM algorithm](#).



# Plan

---

The BOS model for ordinal data

The Latent Block Model

Model inference

Numerical experiments

Model parameter estimation accuracy

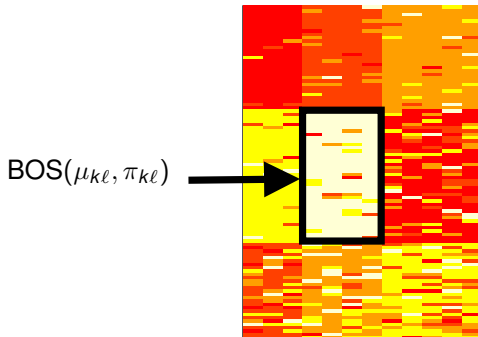
Ability of ICL-BIC to select the number of co-clusters

Comparison with competitors

# Latent Block Model

---

## Latent Block Model (LBM)



# Latent Block Model

---

## Latent Block Model (LBM)

$n \times d$  random variables  $\mathbf{x}$  are assumed to be independent once the row  $\mathbf{v} = (v_{ik})_{i,k}$  and column  $\mathbf{w} = (w_{h\ell})_{h,\ell}$  partitions are fixed:

$$p(\mathbf{x}; \theta) = \sum_{\mathbf{v} \in V} \sum_{\mathbf{w} \in W} p(\mathbf{v}; \theta) p(\mathbf{w}; \theta) p(\mathbf{x} | \mathbf{v}, \mathbf{w}; \theta)$$

with

- $V$  ( $W$ ) set of possible partitions of rows (column) into  $K$  ( $L$ ) groups,
- $p(\mathbf{v}; \theta) = \prod_{ik} \alpha_k^{v_{ik}}$  and  $p(\mathbf{w}; \theta) = \prod_{h\ell} \beta_\ell^{w_{h\ell}}$
- $p(\mathbf{x} | \mathbf{v}, \mathbf{w}; \theta) = \prod_{ihk\ell} p(x_{ih}; \mu_{k\ell}, \pi_{k\ell})^{v_{ik} w_{h\ell}}$  where  $p(x_{ih}; \mu_{k\ell}, \pi_{k\ell}) \sim \text{BOS}(\mu_{k\ell}, \pi_{k\ell})$
- $\theta = (\pi_{k\ell}, \mu_{k\ell}, \alpha_k, \beta_\ell)$

# Plan

---

The BOS model for ordinal data

The Latent Block Model

**Model inference**

Numerical experiments

Model parameter estimation accuracy

Ability of ICL-BIC to select the number of co-clusters

Comparison with competitors

# LBM inference

---

## LBM inference

- The aim is to estimate  $\theta$  by maximizing the observed log-likelihood

$$\ell(\theta; \check{\mathbf{x}}) = \sum_{\hat{\mathbf{x}}} \ln p(\mathbf{x}; \theta). \quad (1)$$

with  $\check{\mathbf{x}}$  is the observed data and  $\hat{\mathbf{x}}$  the unobserved one

- $\mathbf{v}$  and  $\mathbf{w}$  are missing
- EM is not computationally tractable
- $\Rightarrow$  variational or stochastic version should be used

# LBM inference

---

## SEM-Gibbs algorithm for LBM inference

- init :  $\theta^{(0)}$ ,  $\mathbf{w}^{(0)}$
- SE step
  - generate the row partition  $v_{ik}^{(q+1)} | \check{\mathbf{x}}, \mathbf{w}^{(q)}$

$$p(v_{ik} = 1 | \check{\mathbf{x}}, \mathbf{w}^{(q)}; \theta^{(q)}) = \frac{\alpha_k^{(q)} f_k(x_i^{(q)} | \mathbf{w}^{(q)}; \theta^{(q)})}{\sum_{k'} \alpha_{k'}^{(q)} f_{k'}(x_i^{(q)} | \mathbf{w}^{(q)}; \theta^{(q)})}$$

- generate the column partition  $w_{h\ell}^{(q+1)} | \check{\mathbf{x}}, \mathbf{v}^{(q+1)}$

$$p(w_{h\ell} = 1 | \check{\mathbf{x}}, \mathbf{v}^{(q+1)}; \theta^{(q)}) = \frac{\beta_\ell^{(q)} g_\ell(x_{.h}^{(q)} | \mathbf{v}^{(q+1)}; \theta^{(q)})}{\sum_{\ell'} \beta_{\ell'}^{(q)} g_{\ell'}(x_{.h}^{(q)} | \mathbf{v}^{(q+1)}; \theta^{(q)})}$$

- M step
  - Estimate  $\theta$ , conditionally on  $\mathbf{v}^{(q+1)}$ ,  $\mathbf{w}^{(q+1)}$  obtained at the SE step (and also conditionally to  $\check{\mathbf{x}}$ ), using the EM algorithm for BOS inference.

# LBM inference

---

## SEM-Gibbs algorithm for LBM inference

- $\hat{\theta}$  is obtained by mean / mode of the sample distribution (after a burn in period)
- final bipartition  $(\hat{\mathbf{v}}, \hat{\mathbf{w}})$  estimated by MAP conditionally on  $\hat{\theta}$

# LBM inference

---

## Choosing $K$ and $L$

We propose to adapt the ICL-BIC criterion developed in (Keribin et al. 2014) for categorical data coclustering based on the multinomial distribution.

Thus,  $K$  and  $L$  can be chosen by maximizing

$$\text{ICL-BIC}(K, L) = \log p(\mathbf{x}, \hat{\mathbf{v}}, \hat{\mathbf{w}}; \hat{\theta}) - \frac{K-1}{2} \log n - \frac{L-1}{2} \log d - \frac{KL}{2} \log(nd)$$



# LBM inference

---

## Choosing $K$ and $L$

We propose to adapt the ICL-BIC criterion developed in (Keribin et al. 2014) for categorical data coclustering based on the multinomial distribution.

Thus,  $K$  and  $L$  can be chosen by maximizing

$$\text{ICL-BIC}(K, L) = \log p(\mathbf{x}, \hat{\mathbf{v}}, \hat{\mathbf{w}}; \hat{\theta}) - \frac{K-1}{2} \log n - \frac{L-1}{2} \log d - \frac{KL}{2} \log(nd)$$

## Missing data

Within the SEM-Gibbs framework, missing data can be easily taken into account by considering them as an additional missing random variable to be simulated in the SE step.

# Plan

---

The BOS model for ordinal data

The Latent Block Model

Model inference

**Numerical experiments**

Model parameter estimation accuracy

Ability of ICL-BIC to select the number of co-clusters

Comparison with competitors

# Plan

---

The BOS model for ordinal data

The Latent Block Model

Model inference

Numerical experiments

**Model parameter estimation accuracy**

Ability of ICL-BIC to select the number of co-clusters

Comparison with competitors

# Experimental setup

---

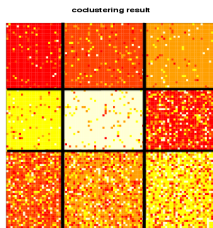
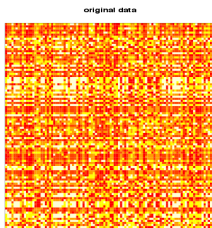
- $K = L = 3$  clusters in row and column
- $d = 100$  ordinal variables with  $m = 5$  levels
- $n = 100$  observations
- values of  $(\mu_{k\ell}, \pi_{k\ell})$

| Setting 1 |         |         |         | Setting 2 |         |         |         |
|-----------|---------|---------|---------|-----------|---------|---------|---------|
| $k/\ell$  | 1       | 2       | 3       | $k/\ell$  | 1       | 2       | 3       |
| 1         | (1,0.9) | (2,0.9) | (3,0.9) | 1         | (1,0.2) | (2,0.2) | (3,0.2) |
| 2         | (4,0.9) | (5,0.9) | (1,0.5) | 2         | (4,0.2) | (5,0.2) | (1,0.1) |
| 3         | (2,0.5) | (3,0.5) | (4,0.5) | 3         | (2,0.1) | (3,0.1) | (4,0.1) |

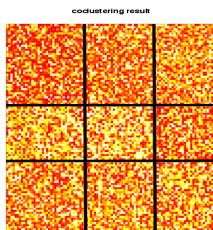
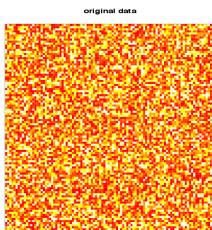
# Example of data

---

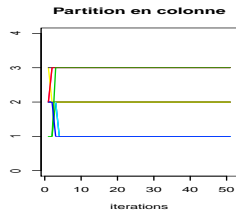
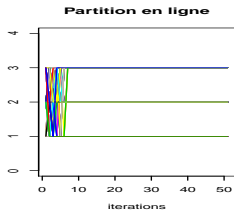
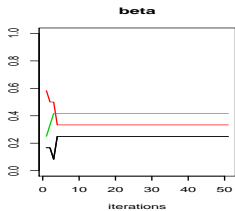
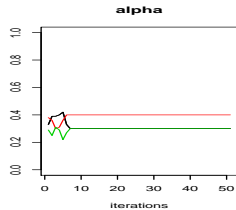
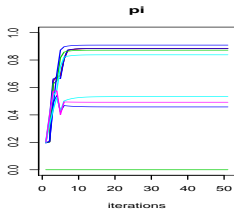
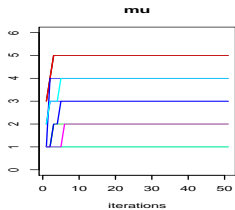
Setting 1:



Setting 2:

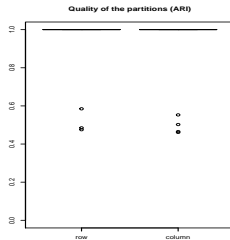
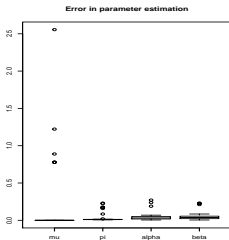


# How many iterations do we need ?

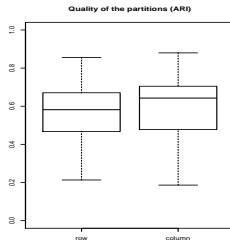
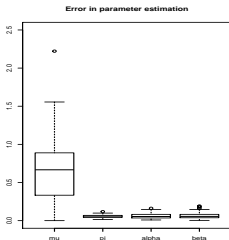


# Accuracy of estimation and co-clustering

Setting 1:



Setting 2:



# Plan

---

The BOS model for ordinal data

The Latent Block Model

Model inference

Numerical experiments

Model parameter estimation accuracy

**Ability of ICL-BIC to select the number of co-clusters**

Comparison with competitors



# Selection of the number of co-clusters

---

## Exp. setup

- experimental setting 1
- ICL-BIC computed for 2 to 4 clusters in row / column
- 50 simulations

## Results

|   |   | L |    |   |
|---|---|---|----|---|
|   |   | 2 | 3  | 4 |
| K | 2 | 0 | 0  | 0 |
|   | 3 | 0 | 46 | 3 |
|   | 4 | 0 | 1  | 0 |

# Plan

---

The BOS model for ordinal data

The Latent Block Model

Model inference

Numerical experiments

Model parameter estimation accuracy

Ability of ICL-BIC to select the number of co-clusters

Comparison with competitors

# Comparison with competitors

---

## Competitors

- R package *blocksluter* for nominal data.
- R package *blocksluter* for continuous data.
- *Optimal* is the Bayes classifier using the true model parameter values.

## Results

|                | setting 1     |               | setting 2     |               |
|----------------|---------------|---------------|---------------|---------------|
|                | ARI row       | ARI column    | ARI row       | ARI column    |
| BOS            | 0.971 (0.117) | 0.960 (0.139) | 0.581 (0.149) | 0.589 (0.171) |
| bc categ.      | 1(0)          | 0(0)          | 0.288 (0.088) | 0.018 (0.055) |
| bc conti.      | 0.841 (0.290) | 0.833 (0.288) | 0.421 (0.103) | 0.270 (0.110) |
| <i>Optimal</i> | 1(0)          | 1(0)          | 0.761 (0.087) | 0.759 (0.087) |

# Conclusions

---

## Results

- BOS: new probability distribution for ordinal data which respects the ordinal scale of data
- (co-)clustering algorithms have been developed based on the BOS model
- Applications are welcome !

# References

---

- [1] Biernacki & Jacques (2015), Model-based clustering of multivariate ordinal data relying on a stochastic binary search algorithm, to appear in *Statistics and Computing*.
- [2] Agresti (2010), *Analysis of ordinal categorical data*, Wiley.
- [3] Govaert, G. and Nadif, M. (2013). *Co-Clustering*. Wiley-ISTE.